*Project Title:* Most Significant correlations as pertaining to ride share trip DATA

*Team Members:* Victor Galstyan, Ruddy Simonpour, Moumita Ghanti, Ghaith Ramahi.

*Project Description/Outline:* We will be combining the Uber data with the weather in that given area. Thereafter, we want to see if there are any correlations between the weather and the rides in the New York city area. In conclusion, we need to use the Uber data retrieved and match it to the given weather on that day and time of the ride. In addition, we need to compare the different time stamps throughout the day, into time buckets, to help further analyze how weather affects the uber rides and the locations they ordered from.

***Research Questions & Answers:***

* Are there any correlations between the weather and the frequency of Uber rides placed? Null Hypothesis: Is there no relation between the ride share ride frequency and the weather?
* How does time in the day affect the frequency of uber rides placed? Null Hypothesis: The time of day has no effect on the ride share ride placed?
* As time progresses, does the frequency of the ride share rides increase? Null Hypothesis: There is no change in the frequency of rides as time progresses.
* How does Time of Day affect Trip Duration of ride share rides? Null Hypothesis: There is no effect on ride share rides throughout the time of the day.
* How does the SR\_Flag, or type of ride share ride, affect the Trip Duration? Null Hypothesis: Trip Duration is not affected by the SR\_Flag, or type of ride share ride

*Datasets to be Used:* Uber rides data set, open weather, google api, New York city weather data set, Uber prices/ride data set.

*Rough Breakdown of Tasks: (TBD)*

* Victor:
* Ruddy:
* Moumita:
* Ghaith:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

First approach: cleaning the data sets. Basically, we are combing the data frames we have to one set; which includes:

* Ride
* Weather
* Location in New York city.
* Time of day

Second approach: Begin analyzing the trends that appear from the combined data frame.

Third approach: Modeling our data through illustrating visualized graphs, and/ or, maps to help further explain our findings and results.

***Points of Correlation:***

Time of Day vs Number of Rides

Wind Speed vs Number of Rides

Humidity vs Number of Rides

Temperature vs Number of Rides

Time of Day vs Trip Duration

Temperature vs Trip Duration

Zip Code vs Trip Duration

Humidity vs Trip Duration

Zip Code vs Number of Rides

Time of Day vs Trip Duration

SR\_Flag vs Trip Duration

Day of Week Vs Trip Duration  
Day of Week Vs Number of Rides